10/567647

IAP20 Rec'd PCT/PTO 09 FEB 2006

DESCRIPTION

APPARATUS AND METHOD FOR DETECTING MOVING OBJECT

Technical Field

[0001] The present invention relates to an apparatus and method for detecting a moving object from a video stream generated by coding a video.

Background Art

[0002] A conventional example of this moving object detection apparatus is described in Patent Document 1.

[0003] This moving object detection apparatus is designed to extract a motion vector used for a motion predictive compensation coding scheme and detect a moving object at a high speed by regarding the motion vector as motion of the object in a certain region, without decoding a video stream. FIG. 1 shows the conventional moving object detection apparatus described in Patent Document 1.

[0004] In FIG. 1, a coding mode of an image block, motion compensation mode and motion vector information decoded by variable-length decoding section 1801 and pattern information detected by pattern information detection section 1802 are sent to moving object detection processing section 1803. Mobile object detection

processing section 1803 decides whether or not this image block is a moving object using the information. This decision is made using the motion vector, spatial similarity decision, temporal similarity decision or the like.

Patent Document 1: Unexamined Japanese Patent Publication No.HEI 10-75457

Disclosure of Invention

Problems to be Solved by the Invention

[0005] However, since the above described conventional configuration only depends on a motion vector which does not always express the motion of the object accurately, it cannot be said to provide highly accuracy. That is, in many cases, a motion vector generation method searches for a reference region where a compression rate of coding is high from images before and after the region being coded and regards the reference to the searched region as the motion vector. For this reason, the accuracy of detection of a moving object using only the motion vector is not high.

[0006] It is an object of the present invention to provide an apparatus and method for detecting a moving object capable of detecting the moving object at a high speed, with high accuracy and low processing load, from a video stream video-coded using a band division method for dividing an image into reduced image, horizontal

direction component, vertical direction component and diagonal direction component and motion predictive compensation coding.

Means for Solving the Problem

[0007] The moving object detection apparatus according to the present invention adopts a configuration having a motion information extraction section that extracts motion information from a video stream video-coded using layered coding whereby a video is coded with being divided into a plurality of layers and motion predictive compensation coding, an edge information extraction section that extracts edge information from the video stream and a moving object detection section that detects a moving object using the motion information and the edge information and outputs the detection result.

[0008] The moving object detection method according to the present invention is a method for detecting a moving object from a video stream, having a step of extracting motion information from a video stream video-coded using layered coding whereby a video is coded with being divided into a plurality of layers and motion predictive compensation coding, a step of extracting edge information from the video stream and a step of detecting a moving object using the extracted motion information and the edge information, the steps being executed by the moving object detection apparatus that detects the

moving object.

Advantageous Effect of the Invention

[0009] According to the present invention, it is possible to detect contours of a moving object at a high speed, with high accuracy and low processing load without decoding the video, from a video stream video-coded using a band division method whereby an image is divided into a reduced image, a horizontal direction component, a vertical direction component and a diagonal direction component and motion predictive compensation coding. Furthermore, it is also possible to decode the video at the same time.

Brief Description of Drawings

[0010]

FIG. 1 illustrates the configuration of a conventional moving object detection apparatus;

FIG. 2 illustrates the configuration of a video decoding apparatus according to Embodiment 1 of the present invention;

FIG. 3 is conceptual diagram of bit plane coding according to Embodiment 1 of the present invention;

FIG. 4 is a flow chart showing the operation of the video decoding apparatus according to Embodiment 1 of the present invention;

FIG. 5 is a flow chart showing the operation of moving

object detection processing by the video decoding apparatus according to Embodiment 1 of the present invention;

FIG. 6 is a stream structural diagram of an expanded layer according to Embodiment 1 of the present invention;

FIG. 7 is a stream structural diagram of bit plane k of the expanded layer according to Embodiment 1 of the present invention;

FIG. 8 is a stream structural diagram of bit plane k of expanded layer j according to Embodiment 1 of the present invention;

FIG. 9 is stream structural diagram of a basic layer according to Embodiment 1 of the present invention;

FIG. 10 is a stream structural diagram of region j of the basic layer according to Embodiment 1 of the present invention;

FIG. 11A shows an example of a horizontal direction component in an 8×8 pixel region according to Embodiment 1 of the present invention;

FIG. 11B shows another example of the horizontal direction component in an 8×8 pixel region according to Embodiment 1 of the present invention;

FIG. 11C is a further example of the horizontal direction component in an 8×8 pixel region according to Embodiment 1 of the present invention;

FIG. 12 shows the configuration of a video monitoring system according to Embodiment 2 of the present invention;

FIG. 13 shows the configuration of an automatic tracking camera according to Embodiment 2 of the present invention;

FIG. 14 shows the configuration of a video coding apparatus according to Embodiment 2 of the present invention;

FIG. 15 is a flow chart showing the operation of the automatic tracking camera according to Embodiment 2 of the present invention;

FIG. 16 is a flow chart showing the operation of the video coding apparatus according to Embodiment 2 of the present invention;

FIG. 17 is a flow chart showing the operation of the video monitoring apparatus according to Embodiment 2 of the present invention;

FIG. 18 is a sequence diagram showing the operation of the video monitoring system according to Embodiment 2 of the present invention;

FIG. 19 shows the configuration of a video decoding apparatus according to Embodiment 3 of the present invention; and

FIG. 20 is a flow chart showing the operation of the video decoding apparatus according to Embodiment 3 of the present invention.

Best Mode for Carrying Out the Invention

[0011] Now, embodiments of the present invention will

be described in detail with reference to the attached drawings below.

[0012]

(Embodiment 1)

Embodiment 1 shows a case where the method and apparatus for detecting a moving object according to the present invention is applied to a video decoding apparatus. That is, Embodiment 1 is designed to be able to decode a video stream and at the same time detect a moving object within a video at a high speed and with high accuracy.

[0013] First, a video stream used in this embodiment will be explained. This video stream consists of a basic layer and an expanded layer, and the basic layer can be decoded singly to obtain a video with low resolution. The expanded layer is additional information capable of improving image quality of the basic layer and obtaining a video with high resolution and includes edge components in horizontal, vertical and diagonal directions (horizontal direction component, vertical direction component and diagonal direction component).

[0014] Next, a method for generating this video stream will be explained.

[0015] First, an input image is band-divided to generate a reduced image, a horizontal component, a vertical component and a diagonal component. Furthermore, the reduced image is coded by motion predictive compensation coding as a basic layer which can singly decode a video.

The horizontal direction component, the vertical direction component and the diagonal direction component are then coded through bit plane coding as an expanded layer to improve image quality of the video obtained by decoding the basic layer.

[0016] Here, the band division will be explained. In the band division, an image is divided into four components; reduced image, horizontal component, vertical component and diagonal component. This band division is performed using a wavelet transform, a combination of high pass filter, low pass filter and downsampler, or the like. Furthermore, the reduced image, horizontal direction component, vertical direction component and diagonal direction component obtained through the band division can be restored to the original image through a band combination. The horizontal direction component, vertical direction component and diagonal direction component obtained through this band division are differences in pixel values from adjacent pixels that can be mathematically calculated, and need not always express contours of an object. For example, in the case of a monochrome horizontal stripe pattern, strong vertical components appear on its color boundary as a horizontal line.

[0017] FIG. 2 is a block diagram showing the configuration of video decoding apparatus 100 according to Embodiment 1 to which the method and apparatus for

detecting a moving object of the present invention are applied.

[0018] In FIG. 2, video decoding apparatus 100 is provided with stream input section 101, basic layer decoding section 102, expanded layer decoding section 103, band combination section 104, video output section 105, moving object detection section 106 and detection result output section 107.

[0019] Note that basic layer decoding section 102, expanded layer decoding section 103 and band combination section 104 correspond to the video decoding section of the present invention, basic layer decoding section 102 corresponds to the motion information extraction section, expanded layer decoding section 103 corresponds to the edge information extraction section and moving object detection section 106 corresponds to the moving object detection section.

[0020] Here, the video decoding section decodes an input video stream, generates and outputs the video. The motion information extraction section extracts motion information from the input video stream and outputs it to the moving object detection section. The edge information extraction section extracts edge information from the input video stream and outputs it to the moving object detection section. The moving object detection section detects a moving object from the input edge information and motion information.

[0021] Next, the operation of video decoding apparatus 100 configured as shown above will be explained.

[0022] FIG. 4 is a flow chart showing the operation of video decoding apparatus 100 according to Embodiment 1 shown in FIG. 2. The operation shown in the flow chart of FIG. 4 may also be made executable by software by causing a CPU (not shown) to execute a control program stored in a storage apparatus (not shown) (e.g., ROM, flash memory or the like).

[0023] First, stream input section 101 receives a video stream from the outside of video decoding apparatus 100 and outputs a basic layer of the video stream to basic layer decoding section 102 and an expanded layer to expanded layer decoding section 103 respectively (step S301).

[0024] Next, basic layer decoding section 102 extracts motion information from the basic layer input from stream input section 101 and outputs it to moving object detection section 106. Furthermore, expanded layer decoding section 103 extracts the edge information from the expanded layer input from stream input section 101 and outputs it to moving object detection section 106. Object detection section 106 then detects a moving object using the motion information and edge information input from basic layer decoding section 102 and expanded layer decoding section 103, generates a moving object detection result and outputs it to detection result output section

107 and band combination section 104 (step S302).

[0025] A video may or may not include a moving object, and when the video includes a moving object, the number of moving objects may be one or plural.

[0026] The moving object detection processing in step S302 will be explained in further detail below.

[0027] FIG. 5 is a flow chart showing an example of steps of the moving object detection processing in FIG. 4.

[0028] First, in step S401, edge information extraction processing is carried out. More specifically, expanded layer decoding section 103 extracts codes including information about the expanded layer up to a specific bit plane input from stream input section 101, generates edge information and outputs it to moving object detection section 106.

[0029] Here, the bit plane coding will be explained.

[0030] This bit plane refers to a bit string with only the same bit positions of several numerical data expressed in binary numbers lined up. A method of coding for each bit plane is called a "bit plane coding" and has excellent performance of adjusting data quality as described in Weiping Li, "Overview of Fine Granularity Scalability in MPEG-4 Video Standard", IEEE Transaction on Circuits and Systems for Video Technology, vol. 11, pp. 301-317, Mar. 2001.

[0031] FIG. 3 shows a concept of bit plane coding and this will be explained as one that expresses a certain

region of a horizontal direction component.

[0032] In FIG. 3, one column expresses 1 pixel of a horizontal component in binary numbers (pixel 1, pixel 2). One row expresses a bit plane in a certain region of the horizontal direction component (bit plane 1, bit plane 2), that is, it is a set of only bits of the same position of each pixel. The higher the position of bit plane, the stronger edge of the horizontal direction component the bit plane can express. Edge information is obtained by coding information about the highest bit plane up to a specific bit plane lined up. For example, the edge information includes information such as the amount of code per bit plane up to a specific bit plane for each region, for example, 8×8 pixels and 16×16 pixels. The horizontal direction component, vertical direction component and diagonal direction component include many "0"s, and therefore when there are many "0"s, bit plane coding is performed so as to shorten the code length. Therefore, the more "1"s are included, the longer the code length of the bit plane of the region of each of the horizontal direction component, vertical direction component and diagonal direction component becomes.

[0033] FIG. 6 shows a data structure of an expanded layer of this embodiment. The expanded layer shown in FIG. 6 is a code corresponding to one image and includes information about n bit planes and m regions. The expanded layer corresponding to one image stores image

header information 501 and information 502 on bit plane 1 which indicates the highest bit plane to bit plane n which is the lowest bit plane.

[0034] FIG. 7 shows a data structure of bit plane k of the expanded layer in FIG. 6 and bit plane k of the expanded layer includes bit plane header information 601 and code 602 of bit planes k of region 1 to region m.

[0035] FIG. 8 shows the data structure of bit plane k of region j of the expanded layer in FIG. 7 and bit plane k of region j of the expanded layer includes code 701 of the pixel component of the corresponding region and termination signal 702 indicating that the region code is terminated.

[0036] With the above described data structure, it is just possible to extract bit plane information by sequentially searching for termination signals of those regions from the highest bit plane to a specific bit plane within a video stream and counting the code length between the region termination signals. Thus, expanded layer decoding section 103 can generate edge information at a high speed.

[0037] Next, in step S402, motion information extraction processing is performed. More specifically, basic layer decoding section 102 extracts information about the motion vector from the basic layer input from stream input section 101, generates motion information and outputs it to moving object detection section 106.

[0038] This motion information is used for motion predictive compensation of the basic layer and includes information about whether it is for each region that motion predictive compensation coding or in-frame coding is perfomed, information about magnitude and direction of the motion vector and the image referenced by the motion vector, information about whether it is the entire image that motion predictive compensation coding or in-frame coding is performed on, or the like.

[0039] FIG. 9 shows the data structure of the basic layer of this embodiment. The basic layer shown in FIG. 9 is a code corresponding to one image and includes information about m regions. That is, the one-image basic layer includes image header information 801 and information 802 on region 1 to region m. FIG. 10 shows the data structure of region p of the basic layer in FIG. 9 and region p of the basic layer includes region header information 901, motion vector 902, pixel component code 903 and termination signal 904 indicating that the region code is terminated.

[0040] A motion vector just can be extracted by searching for header information 901 and termination signal 904 of those regions from the video stream and decoding only motion vector 902 located at a fixed position from that position. This allows basic layer decoding section 102 to generate motion information at a high speed.

[0041] In step S403, processing of detecting contours

of a moving object is performed. More specifically, moving object detection section 106 detects a region of contours of the moving object using motion information and edge information input from basic layer decoding section 102 and expanded layer decoding section 103 and stores the result in moving object detection section 106.

[0042] Here, the method for detecting the contour region will be explained.

[0043] That is, suppose condition 1 is that the code length calculated from the bit planes of a horizontal direction component, vertical direction component and diagonal direction component corresponding to a certain region, for example, the total code length of the respective amounts of code from the highest bit plane to the third bit plane should be equal to or greater than threshold A. Note that this threshold A is a reference value whereby an edge is decided to be a weak edge.

[0044] Furthermore, suppose condition 2 is that the total code length of the above described regions should be equal to or smaller than threshold B. This threshold B is a reference value to identify an image which is not an edge such as a stripe pattern.

[0045] It is then decided whether the edge information including the region indicates a dot, line or plane or not, and when the total code length of the above described regions satisfies these condition 1 and condition 2, it is decided to be a line appearing on the contours of an

object. A specific example will be explained using FIG. 11 below.

[0046] FIG. 11A to FIG. 11C show examples of a horizontal direction component in an 8×8 pixel region. For simplicity of explanation, pixel values are expressed by binary numbers and cells including "1" from the highest bit plane to a specific bit plane are shown in black and cells not including "1" are shown in white. FIG. 11A shows a horizontal direction component when noise and small points or the like exist within the region, FIG. 11B shows a horizontal direction component where a vertical line exists within the region and FIG. 11C shows a horizontal direction component when the entire region is part of, for example, a stripe pattern. When the regions expressed in FIG. 11A to FIG. 11C are coded, the amount of code increases in order of FIG. 11A, FIG. 11B, and FIG. 11C according to the number of values other than 0 included in each region. The same applies to the vertical direction component and diagonal direction component. At this time, assuming that threshold A is 8 and threshold B is 32, the region shown in FIG. 11B in which a relationship of "threshold A < the above described total value < threshold B" holds can be decided to include lines appearing in the contours of an object. Here, threshold A < threshold B.

[0047] Furthermore, as more simple contour extraction, the region where a relationship of "threshold A < the

above described total value" holds can also be decided to include lines appearing in the contours of the object by using only threshold A.

[0048] Furthermore, whether a certain region decided to be contours is the contours of the moving object or not is determined by whether the region satisfies condition 3 or condition 4 below or not.

[0049] That is, because condition 3 requires that the magnitude of the motion vector of the region be smaller than threshold C and motion of the target moving object needs to show motion to a certain degree or higher.

[0050] Condition 4 requires that the magnitude of a vector corresponding to the difference between a motion vector of a region and a surrounding motion vector be smaller than threshold D. This decides whether or not the moving object performs the same motion as that of the surrounding region. The number of surrounding motion vectors need not be one. Condition 4 in such a case will be explained. First, a plurality of surrounding motion vectors are extracted and the magnitude of the vector corresponding to the difference from the motion vector of the region is determined for each surrounding motion vector. Condition 4 in this case requires that the total value of the difference vectors be smaller than threshold D.

[0051] The following conditions other than that described above can also be considered for condition 4.

For example, when a plurality of motion vectors are selected as surrounding motion vectors, the sum total of the sum of squares of the difference between the X-direction components (horizontal direction components) of the motion vector in the region and surrounding region and the sum of squares of the difference between the Y-direction components (vertical direction components) (hereinafter referred to as "variance") can also be used as a reference. Condition 4 in this case requires that the above described variance be smaller than threshold D. When condition 4 is satisfied, the motion vector of the region is considered to have the same direction and magnitude as those of the surrounding regions and the object is decided not to be a moving object. Furthermore, the calculation of a variance is not limited to this and the variance may also be calculated as the sum total of the products of the absolute values of the difference in the magnitude of the motion vector and the absolute value of the difference in the angle in surrounding regions. Any method can be adopted if it at least makes it possible to decide whether the motion vector in the region has the direction and magnitude different from those of the surrounding motion vectors or not.

[0052] When this condition 3 or condition 4 is satisfied, the region is decided not to be the region of the moving object. In the case of a frame including no motion vector such that the overall image is in-frame

coded, a frame including a motion vector is waited without deciding contours. This is because it is not possible to detect any motion from a frame with no motion vector.

[0053] Mobile object detection section 106 decides that regions satisfying condition 3 or condition 4 out of the regions determined to be contours of an object from the above described condition 1 and condition 2 are not the contours of the moving object. This is because the contours of a moving object move at a speed different from that of surroundings.

[0054] Next, in step S404, processing of detecting the inside of the moving object is performed. More specifically, moving object detection section 106 detects the region inside the moving object using motion information input from basic layer decoding section 102 and the stored detection result of contours of the moving object. The detection result of the internal region is stored in moving object detection section 106.

[0055] Here, the method for detecting the internal region will be explained below.

[0056] That is, the condition whereby a certain region is decided to be the inside of a moving object is to satisfy condition 5 or condition 6 shown below.

[0057] Condition 5 requires that the region be in the neighborhood of the region decided to be the contours or the inside of the moving object and that a variance in the magnitude and direction of the motion vector with

respect to the neighboring regions be smaller than threshold E, where threshold E is a reference value when the contours and inside of the moving object are decided to move at the same speed.

[0058] Condition 6 requires that the region be surrounded by the region decided to be the contours or the inside of the moving object and this is because the inside of the moving object is surrounded by the contours.

[0059] Next, in step S405, processing of removing error detection of the moving object is performed. More specifically, moving object detection section 106 removes an erroneously detected region from the stored detection results of the contours of and the region inside the moving object, generates a moving object detection result and outputs it to detection result output section 107 and band combination section 104.

[0060] A decision condition for this erroneously detected region is that there are a few regions decided to be the contours or the inside of the moving object in the surroundings and this is because when a too small moving object is detected, the possibility of erroneous detection is high.

[0061] Mobile object detection section 106 generates a moving object detection result from the region of the moving object obtained as shown above. The moving object detection result is, for example, as shown below.

[0062] First, it is information describing for each

region whether or not it is a region of the moving object and second, information defining one rectangle or ellipse circumscribing one moving object and describing coordinates and size for each rectangle or ellipse.

[0063] When information about the inside of the moving object is not necessary, the processing of detecting the inside may be omitted.

[0064] Furthermore, the method for detecting the moving object is not limited to the moving object detection method using a motion vector, but other methods can also be used if combined with the edge information of the present invention.

[0065] According to the moving object detection method in this embodiment, if the basic layer includes a motion vector and the expanded layer at least includes codes up to the bit plane of a certain bit position, it is possible to detect a moving object at a high speed, with high accuracy and low processing load even when transmission is performed at a low bit rate and the image quality is poor.

[0066] Next, in step S303, the result of detecting the moving object is output. More specifically, detection result output section 107 outputs coordinates of the region of the moving object input from moving object detection section 106 to the outside.

[0067] Next, in step S304, basic layer decoding processing is performed. More specifically, basic layer

decoding section 102 subjects the basic layer of the video stream input from stream input section 101 to motion predictive compensation decoding, generates a reduced image and outputs it to band combination section 104.

[0068] Next, in step S305, expanded layer decoding processing is performed. More specifically, expanded layer decoding section 103 subjects the expanded layer of the video stream input from stream input section 101 to bit plane decoding, generates a horizontal direction component, vertical direction component and diagonal direction component and outputs the components to band combination section 104.

[0069] Next, in step S306, band combination processing is performed. More specifically, band combination section 104 band-combines the reduced image input from basic layer decoding section 102 and the horizontal direction component, vertical direction component and diagonal direction component input from expanded layer decoding section 103, generates a decoded image and outputs the decoded image to video output section 105. Furthermore, band combination section 104 may also emphasize the region including the moving object of the decoded image using the moving object detection result input from moving object detection section 106.

[0070] Here, the emphasis of the region of this moving object will be explained. For example, band combination section 104 colors a decoded video of only the region

of the moving object or performs processing such as enclosing the region of the moving object with a frame or the like. Furthermore, it is also possible to set all pixel values of the reduced image obtained by decoding the basic layer to "0" to band-combine and generate an image made up of only contours, and further emphasize the region of the moving object.

[0071] By doing so, only the moving object becomes quite noticeable in the video made up of contours and it is easier for a supervisor who monitors a plurality of monitoring videos simultaneously to detect an abnormality or suspicious figure. Furthermore, when the bit rate of the basic layer is very low due to restrictions on the communication speed and only videos of extremely bad image quality can be generated, contours alone may rather help recognition of details. Or in an environment in which a processing capacity is limited, for example, when a plurality of camera videos are displayed, displaying only contours may make it easier to monitor important regions with low processing load.

[0072] Next, in step S307, video output processing is performed. More specifically, video output section 105 outputs the decoded video input from band combination section 104 to the outside.

[0073] Note that it is possible to only detect the moving object without carrying out decoding processing. At this time, the video cannot be obtained, but since processing

from basic layer decoding processing (step S304) to video output processing (step S307) are not performed, it is possible to detect a moving object at a higher speed and with lower processing load.

[0074] Next, in step S308, end decision processing is performed. When stream input section 101 decides the presence or absence of the next video stream, for example, and then, if video decoding apparatus 100 needs to perform neither detection of moving objects any longer nor decoding videos, the video decoding apparatus ends the processing, or returns to step S301 otherwise.

[0075] In the foregoing explanations, basic layer decoding processing (step S304) to video output processing (step S307) are performed after the moving object detection processing (step S302 and step S303), but the present invention is not limited to this, it is possible to perform moving object detection processing concurrently with the decoding processing of the basic layer and expanded layer.

[0076] Furthermore, as a method of generating a video stream according to another coding method using band division, it is possible to use a method of performing band division after motion predictive compensation on an input image and then bit plane coding. However, according to this method, even when an image for which the difference between preceding and following images is taken through motion predictive compensation is

band-divided, it is not possible to obtain the horizontal direction component, vertical direction component and diagonal direction component which are generated on contours of the object. In this case, only the horizontal direction component, vertical direction component and diagonal direction component of the image, entirety of which is in-frame coded are used.

[0077] Furthermore, the expanded layer may include not only the horizontal direction component, vertical direction component and diagonal direction component but also information corresponding to the difference of the images obtained by decoding the reduced image and basic layer.

[0078] As shown above, Embodiment 1 provides a section that extracts edge information and motion information from information about a horizontal direction component, vertical direction component and diagonal direction component obtained by directly band-dividing an input image and a video stream including a motion vector generated through motion predictive compensation, and therefore, it is possible to detect a moving object at a high speed, with high accuracy and low processing load without decoding the video stream made up of a basic layer using motion predictive coding and an expanded layer using bit plane coding of the horizontal direction component, vertical direction component and diagonal direction component.

[0079] Furthermore, according to Embodiment 1, it is possible to extract motion information from a video stream of the basic layer and extract the edge information from a video stream of the expanded layer, and when the motion information indicates that there is no motion, it is possible to stop processing such as extraction of edge information and alleviate processing load, and when the edge information indicates that there is no edge, it is possible to stop processing such as extraction of motion information and alleviate processing load, thus enabling the detection of contours of the object at a high speed. At this time, any of the motion information and edge information can be extracted first or motion information and edge information can be extracted concurrently.

[0080] Furthermore, according to Embodiment 1, it is possible to detect a moving object with only a motion vector and edge information of some bit planes and thereby detect the moving object at a high speed and with high efficiency even from a low bit rate video stream in a situation in which the communication speed is restricted.

[0081] Furthermore, according to Embodiment 1, expanded layer decoding section 103 extracts edge information necessary to detect a moving object and basic layer decoding section 102 extracts motion information, and therefore, the video decoding processing and moving object detection processing can share some section and processing, and it is possible to perform detection of

a moving object and video decoding simultaneously and at a high speed, and reduce the overall scale of the apparatus.

[0082] Furthermore, according to Embodiment 1, expanded layer decoding section 103 can generate edge information at a high speed just by searching for a start signal included in bit plane header 601 within a video stream and termination signal 702 for each region of 8×8 pixels or the like and counting the code length between identification signals.

[0083] Furthermore, according to Embodiment 1, basic layer decoding section 102 just searches for an identification signal for each region of 8×8 pixels or the like within a video stream and decodes a motion vector at a predetermined position from the identification signal, and therefore, it is possible to generate motion information at a high speed.

[0084] Furthermore, according to Embodiment 1, moving object detection section 106 detects contours of a moving object using edge information and motion information, detects the inside of the moving object using the motion information and already detected result and removes erroneous detection, and therefore, it is possible to detect the moving object with high accuracy.

[0085] Furthermore, according to Embodiment 1, band combination section 104 emphasizes the region of a moving object of a decoded video or uses a line drawing in which

a reduced video which is a decoded basic layer is not band-combined, and therefore, it is possible to help a supervisor detect the detection result of the moving object.

[0086]

(Embodiment 2)

Embodiment 2 is a case where the method and apparatus for detecting a moving object according to the present invention is applied to a video monitoring system. The video monitoring system includes an automatic tracking camera provided with a video coding apparatus. That is, the video monitoring system codes a video and generates a video stream, and at the same time, detects a moving object which exists in the video at a high speed, with high accuracy and low processing load, and based on the detection result, it is possible for the automatic tracking camera to automatically track the moving object and perform video monitoring efficiently.

[0087] This video monitoring system will be explained more specifically below.

[0088] FIG. 12 shows the configuration of a video monitoring system according to Embodiment 2 to which the method and apparatus for detecting a moving object of the present invention is applied.

[0089] This video monitoring system includes video monitoring apparatus 1100, communication network 1110 and N automatic tracking cameras 1121 to 112N. The

automatic tracking camera corresponds to the image pickup apparatus of the present invention.

[0090] FIG. 13 is a block diagram showing the configuration of automatic tracking cameras 1121 to 112N according to Embodiment 2. The automatic tracking camera shown in FIG. 13 corresponds to automatic tracking camera 1121 in the video monitoring system shown in FIG. 12.

[0091] In FIG. 13, automatic tracking camera 1121 includes image pickup section 1201, video coding section 1202 and image pickup control section 1203. Other automatic tracking cameras 1122 to 112N also have similar configurations.

[0092] Image pickup section 1201 corresponds to the image pickup section of the present invention and image pickup control section 1203 corresponds to the image pickup control section of the present invention.

[0093] Here, image pickup section 1201 carries out an image pickup function operation such as pan/tilt/zoom and outputs a video captured to video coding section 1202.

[0094] Video coding section 1202 band-divides the input video, generates a video stream including information about the horizontal direction component, vertical direction component and diagonal direction component and motion vector generated by motion predictive compensation.

[0095] Image pickup control section 1203 receives information about a tracking target and a result of moving

object detection, generates and outputs a control signal for carrying out a pan/tilt/zoom for image pickup section 1201.

[0096] FIG. 14 is a block diagram showing the configuration of video coding apparatus 1202 and corresponds to a video coding apparatus to which the method and apparatus for detecting a moving object of the present invention is applied.

[0097] In FIG. 14, video coding section 1202 includes video input section 1301, band division section 1302, basic layer coding section 1303, expanded layer coding section 1304, stream output section 1305, moving object detection section 1306 and detection result output section 1307.

[0098] Note that band division section 1302, basic layer coding section 1303 and expanded layer coding section 1304 correspond to the video coding section of the present invention, and basic layer coding section 1303 corresponds to the motion information extraction section, expanded layer coding section 1304 corresponds to the edge information extraction section and moving object detection section 1306 corresponds to the moving object detection section.

[0099] Here, the video coding section codes an input video, and generates and outputs a video stream. Band division section 1302 that constitutes this video coding section band-divides the input image to generate a reduced

image, horizontal component, vertical component and diagonal component and subjects the reduced image to motion predictive compensation coding to code it as a basic layer capable of singly decoding the video. Furthermore, band division section 1302 subjects these horizontal direction component, vertical direction component and diagonal direction component to bit plane coding and codes them as an expanded layer. Basic layer coding section 1303 extracts motion information from the generated video stream and outputs it to moving object detection section 1306. Expanded layer coding section 1304 extracts edge information from the generated video stream and outputs it to moving object detection section 1306. Mobile object detection section 1306 detects a moving object from the input edge information and motion information. Stream output section 1305 and detection result output section 1307 correspond to the output section of the present invention.

[0100] Next, the operation of automatic tracking camera 1121 according to this embodiment will be explained. FIG. 15 is a flow chart showing the operation of automatic tracking camera 1121 shown in FIG. 13. The flow chart shown in FIG. 15 may also be made executable by software by causing a CPU (not shown) to execute a control program stored in a storage apparatus (not shown) (e.g., ROM, flash memory or the like).

[0101] First, in step S1401, image pickup processing is

performed. More specifically, image pickup section 1201 captures a video which is a monitoring target and outputs the input image to video input section 1301 of video coding section 1202. Furthermore, image pickup section 1201 outputs information about a pan/tilt/zoom and installation location to detection result output section 1307 of video coding section 1202.

[0102] Next, in step S1402, video coding processing is performed. Video coding section 1202 codes an input video input from image pickup section 1202 to generate a video stream and at the same time detects a moving object to generate a moving object detection result. These generated video stream and moving object detection result are output to reception section 1101 of video monitoring apparatus 1100 via communication network 1110. Furthermore, the moving object detection result is output to image pickup control section 1203.

[0103] Next, in step S1403, image pickup control processing is performed. More specifically, image pickup control section 1203 generates a pan/tilt/zoom control signal according to a target tracking command input from camera group control section 1102 of video monitoring apparatus 1100 via communication network 1100 and moving object detection result input from the video coding section and outputs it to image pickup section 1201. Image pickup section 1201 carries out a pan/tilt/zoom based on the control signal input from image

pickup control section 1203.

[0104] Here, this control signal will be explained. When the target tracking command generated by video monitoring apparatus 1100 which will be described later specifies, for example, coordinates and magnification or the like for taking images of a suspicious figure to be captured, image pickup control section 1203 generates a control signal to carry out a pan/tilt/zoom accordingly. When there is a difference between the coordinates to take images of the suspicious figure to be captured and coordinates of the region of the moving object shown in the moving object detection result, image pickup control section 1203 may also correct the difference and generate a control signal. Furthermore, it is also possible to pan the camera such that the moving object to be tracked always occupies a fixed area with respect to the screen. When there is no target tracking command, yet there is a moving object detection result, images are taken with the moving object set as the center of the video. Furthermore, it is also possible to generate a control signal so that all of the plurality of moving objects are accommodated in the video. In addition, especially when there is neither target tracking command nor moving object detection result, it is possible to generate a control signal to cause image pickup section 1201 to perform oscillating motion for the purpose of taking images over a wide range.

[0105] Next, in step S1404, when there is no more need to carry out video monitoring, for example, when power to automatic tracking camera 1121 is turned OFF, the automatic tracking camera 1121 ends its operation or returns to step S1401 otherwise.

[0106] Here, the video coding processing in step S1402 in FIG. 15 will be explained in detail.

[0107] FIG. 16 is a flow chart showing the operation of video coding section 120. The operation shown in the flow chart of FIG. 16 may also be executed by software by causing a CPU (not shown) to execute a control program stored in a storage apparatus (e.g., ROM, flash memory or the like) (not shown).

[0108] First, in step S1501, video input processing is performed. More specifically, video input section 1301 receives an input image from image pickup section 1201 of automatic tracking camera 1121 and outputs it to band division section 1302.

[0109] Next, in step S1502, band division processing is performed. More specifically, band division section 1302 band-divides the input image input from video input section 1301 to generate a reduced image, horizontal direction component, vertical direction component and diagonal direction component, outputs the reduced image to basic layer coding section 1303 and outputs the horizontal direction component, vertical direction component and diagonal direction component to expanded

layer coding section 1304.

[0110] Next, in step S1503, basic layer coding processing is performed. More specifically, basic layer coding section 1303 subjects the reduced image input from band division section 1302 to motion predictive compensation coding to generate a basic layer and outputs it to stream output section 1305. Furthermore, motion information obtained during motion predictive compensation is output to moving object detection section 1306.

[0111] Next, in step S1504, expanded layer coding processing is performed. More specifically, expanded layer coding section 1304 subjects the horizontal direction component, vertical direction component and diagonal direction component input from band division section 1302 to bit plane coding to generate an expanded layer and outputs it to stream output section 1305. Furthermore, edge information obtained during bit plane coding is output to moving object detection section 1306.

[0112] Next, in step S1505, stream output processing is performed. More specifically, stream output section 1305 outputs the basic layer input from basic layer coding section 1303 and the expanded layer input from expanded layer coding section 1304 to reception section 1101 of video monitoring apparatus 1100 via communication network 1110.

[0113] Next, in step S1506, moving object detection processing is performed. More specifically, moving

object detection section 1306 detects a moving object using the motion information input from basic layer coding section 1303 and edge information input from expanded layer coding section 1304, generates a moving object detection result and outputs it to detection result output section 1307.

[0114] The method for detecting a moving object is similar to that of Embodiment 1, and therefore detailed explanations thereof will be omitted here.

[0115] Next, in step S1507, detection result output processing is performed. More specifically, detection result output section 1307 outputs the moving object detection result input from moving object detection section 1306 and information about the pan/tilt/zoom and installation location or the like input from image pickup section 1201 of automatic tracking camera 1121 to reception section 1101 of video monitoring apparatus 1100 via communication network 1110.

[0116] As in the case of the video decoding apparatus described in Embodiment 1, this embodiment can also use other band division methods if it is at least possible to generate a video stream including information about the horizontal direction component, vertical direction component and diagonal direction component and a motion vector generated through motion predictive compensation.

[0117] Next, the configuration of video monitoring apparatus 1100 according to this embodiment will be

explained below.

[0118] In FIG. 12, video monitoring apparatus 1100 is provided with reception section 1101, image recognition section 1102 and camera group control section 1103.

[0119] Image recognition section 1102 corresponds to the image recognition section of the present invention, receives a video stream and moving object detection result, carries out detailed image recognition and outputs the image recognition result to camera group control section 1103.

[0120] Camera group control section 1103 corresponds to the camera group control section of the present invention, receives the image recognition result, and generates and outputs information about the tracking target to cameras 1121 to 112N.

[0121] Next, the operation of video monitoring apparatus 1100 configured as shown above will be explained.

[0122] FIG. 17 is a flow chart showing the operation of video monitoring apparatus 1100.

[0123] First, in step S1601, reception processing is performed. More specifically, reception section 1101 receives the video stream and moving object detection result from automatic tracking camera 1121 via communication network 1110 and outputs them to image recognition section 1102.

[0124] Next, in step S1602, image recognition processing is performed. More specifically, image recognition

section 1102 decodes the video stream using the video stream and moving object detection result input from reception section 1101, performs such as detection or authentication of a figure, face or object using various publicly known image recognition methods, generates the result and outputs them to camera group control section 1103. Furthermore, image recognition section 1102 can further enhance the processing speed by preventing image recognition on any regions other than the region of the moving object included in the moving object detection result.

[0125] Next, in step S1603, camera control processing is performed. More specifically, camera group control section 1103 generates a target tracking command for automatic tracking camera 1121 by using the image recognition result input from image recognition section 1102, and outputs it to image pickup control section 1203 of automatic tracking camera 1121 via communication network 1110. Furthermore, when new tracking of other automatic tracking cameras 1122 to 112N needs to be performed depending on the image recognition result for automatic tracking camera 1121, a new target tracking command is generated and output to image pickup section 1203 of corresponding automatic tracking cameras 1122 to 112N via communication network 1110.

[0126] Here, the target tracking command will be explained.

[0127] When the image recognition result input from image recognition section 1102 indicates, for example, the presence of a suspicious figure in the video, camera group control section 1103 generates a target tracking command including coordinates and magnification or the like to take zoomed images of the suspicious figure. Furthermore, when the suspicious figure exists in the video, yet automatic tracking camera 1121 cannot take any image of the face of the suspicious figure, camera group control section 1103 generates a target tracking command to instruct automatic tracking camera 1122 to take an image of the suspicious figure and generates a target tracking command to instruct automatic tracking camera 1121 to take an image over a wide range including the suspicious figure.

[0128] Next, in step S1604, an end decision is made and if video monitoring need not be performed when, for example, the power to video monitoring apparatus 1100 is turned OFF, video monitoring apparatus 1100 ends the processing or returns to step S1601 otherwise.

[0129] The operation of the video monitoring system configured as shown above will be explained below.

[0130] FIG. 18 is a sequence diagram showing the operation of the video monitoring system according to this embodiment.

[0131] First, automatic tracking camera 1121 takes an image of a monitoring target, generates a video stream

including information about the horizontal direction component, vertical direction component and diagonal direction component and a motion vector generated through motion predictive compensation, obtains a moving object detection result and sends them to video monitoring apparatus 1100 via communication network 1110 (step S1701).

[0132] Video monitoring apparatus 1100 decodes the received video stream and recognizes the target object using the information about the moving object detection result. Video monitoring apparatus 1100 then sends a target tracking command for tracking the target object to automatic tracking camera (step S1702).

[0133] Upon reception of this command, automatic tracking camera 1121 controls the image pickup section and tracks the target. Automatic tracking camera 1121 then sends the video stream or the like at this time to video monitoring apparatus 1100 (step S1703).

[0134] Hereafter, step S1702 and step S1703 are repeated. The video stream or the like from automatic tracking camera 1121 is always sent to video monitoring apparatus 1100 regardless of the presence or absence of a command from video monitoring apparatus 1100.

[0135] As described above, in order to send a video from the automatic tracking camera to the video monitoring apparatus via the communication network, the video monitoring system according to this embodiment needs to

code the video and create a video stream with compressed data. At this time, in the process of generating a video stream, the present invention can detect a moving object simultaneously and report the result information to the video monitoring apparatus, and therefore the video monitoring apparatus need no longer detect the moving object from the received video stream again. This alleviates the processing by the video monitoring apparatus.

[0136] Furthermore, according to Embodiment 2, in the video monitoring system whereby an image captured by an automatic tracking camera at a remote place is received and the video monitoring apparatus performs monitoring and tracking of the video, the automatic tracking camera can share some section and processing, perform processing of video coding into a video stream including information about the horizontal direction component, vertical direction component and diagonal direction component of the image captured and a motion vector generated through motion predictive compensation and moving object detection processing, and can thereby perform accurate detection of a moving object and video coding simultaneously and at a high speed and also reduce the overall scale of the system.

[0137] Furthermore, according to Embodiment 2, the automatic tracking camera can control the image pickup function of a pan/tilt/zoom according to a command from

the video monitoring apparatus determined based on the detection result of the moving object, and therefore, it is possible to efficiently monitor a moving object or suspicious figure or the like.

[0138] Furthermore, according to Embodiment 2, the video monitoring apparatus recognizes the image of only the region of the moving object based on the detection result of the moving object input together with the above described video stream, and therefore, it is possible to alleviate the load of image recognition processing and improve the accuracy of image recognition. Furthermore, this makes it possible to create a video monitoring system capable of controlling more automatic tracking cameras and performing more efficient monitoring.

[0139]

(Embodiment 3)

Embodiment 3 is a method and apparatus for detecting a moving object according to the present invention.

[0140] Of a video stream made up of a basic layer and expanded layer as in the case of Embodiment 1, this embodiment will describe a method for detecting a moving object using only a video stream of the expanded layer. In the video stream of the expanded layer discussed in this embodiment, suppose motion vector information is included at the start of a frame of the video stream of the expanded layer as FGST (FGS Temporal Scalability)

of MPEG-4 FGS (Fine Granularity Scalable coding) defined in ISO/IEC 14496-2 Amendment 2.

[0141] FIG. 19 is a block diagram showing the configuration of moving object detection apparatus 1900 according to Embodiment 3 to which the method and apparatus for detecting a moving object of the present invention is applied.

[0142] In FIG. 19, moving object detection apparatus 1900 is provided with stream input section 1901, motion information extraction section 1902, edge information extraction section 1903, moving object detection section 1904 and detection result output section 1905.

[0143] Unlike Embodiment 1, in this embodiment, stream input section 1901 receives only a video stream of an expanded layer.

[0144] Motion information extraction section 1902 corresponds to the motion information extraction section, edge information extraction section 1903 corresponds to the edge information extraction section and moving object detection section 1904 corresponds to the moving object detection section.

[0145] Here, the motion information extraction section extracts motion information from the video stream of the input expanded layer and outputs it to the moving object detection section. The edge information extraction section extracts edge information from the video stream of the input expanded layer and outputs it to the moving

object detection section. The moving object detection section detects a moving object from the input edge information and motion information.

[0146] Next, the operation of moving object apparatus 1900 configured as described above will be explained below.

[0147] FIG. 20 is a flow chart showing the operation of moving object apparatus 1900 according to Embodiment 3 shown in FIG. 19. The operation shown in flow chart of FIG. 20 is also made executable by software by causing a CPU (not shown) to execute a control program stored in a storage apparatus (not shown) (e.g., ROM, flash memory or the like).

[0148] First, stream input section 1901 receives a video stream of an expanded layer from the outside of moving object detection apparatus 1900 and outputs it to motion information extraction section 1902 and edge information extraction section 1903 (step S2001).

[0149] Next, motion information extraction section 1902 extracts motion information from the expanded layer input from stream input section 1901 and outputs it to moving object detection section 1904 (step S2002).

[0150] Next, edge information extraction section 1903 extracts edge information from the expanded layer input from stream input section 1902 and outputs it to moving object detection section 1904 (step S2003).

[0151] Here, according to FGST defined in MPEG-4 FGS,

motion vectors of the entire region of a frame are stored at the start of the expanded layer of one frame and information about the bit plane is stored following this. Therefore, stream input section 1901 may also input up to the video stream of the motion vector, motion information extraction section 1902 may generate motion information, and stream input section 1901 may input the video stream of the bit plane only when there is motion in the frame and output it to edge information extraction section 1903. Thereby, when there is no motion in the frame, it is possible to omit stream input processing, edge extraction processing and moving object detection processing, and reduce the processing load.

[0152] Next, moving object detection section 1904 detects a moving object using the motion information input from motion information extraction section 1902 and edge information input from edge information extraction section 1903, generates a moving object detection result as in the case of Embodiment 1 and outputs it to detection result output section 1905 (step S2004 to step S2006).

[0153] Next, the result of the moving object detection is output. More specifically, detection result output section 1905 outputs coordinates of the region of the moving object input from moving object detection section 1904 to the outside (step S2007).

[0154] Next, end decision processing is performed. Stream input section 1901 decides the presence or absence

of a subsequent video stream, for example, and then, if moving object detection apparatus 1900 need to perform no more detection of moving objects, moving object detection apparatus 1900 ends the processing or returns to step S2001 otherwise (step S2008).

[0155] As described above, according to Embodiment 3, only a video stream of an expanded layer is input, motion information extraction section 1902 extracts motion information, edge information extraction section 1903 extracts edge information, and therefore it is possible to detect contours of the object at a high speed and with fewer video streams.

[0156] The moving object detection apparatus according to the present invention adopts a configuration including a motion information extraction section that extracts motion information from a video stream video-coded using layered coding whereby a video is coded with being divided into a plurality of layers and motion predictive compensation coding, an edge information extraction section that extracts edge information from the video stream and a moving object detection section that detects a moving object using the motion information and the edge information and outputs the detection result.

[0157] According to this configuration, it is possible to detect object contours without decoding any video stream, detect a moving object from motion information and detect a moving object at a high speed, with high

accuracy and low processing load.

[0158] Furthermore, in the moving object detection apparatus according to the present invention, the edge information extraction section extracts bit plane information from a highest bit plane to Nth (N; natural number) bit plane out of bit plane information obtained by subjecting an image to bit plane coding as edge information from the video stream.

[0159] According to this configuration, by extracting information up to a specific bit plane, it is possible to detect an edge of specific intensity or greater and to therefore detect contours of an object at a high speed. Furthermore, it is possible to detect contours of the object using only a bit plane equal to or higher than a specific bit position without requiring bit planes lower than the specific bit position and realize high accuracy detection at a low bit rate even when a video stream is received via a communication network at a low communication speed.

[0160] Furthermore, in the moving object detection apparatus according to the present invention, the video stream is divided into a plurality of regions, the moving object detection section decides, when the total code length of bit plane information inside the region is equal to or greater than a predetermined first value, that the region is a contour region of the moving object.

[0161] According to this configuration, it is possible

to decide the number of edges which exist inside the region only by confirming the amount of code of bit planes up to a threshold bit position in a certain region of the image and detect the object contours at a high speed.

[0162] Furthermore, in the moving object detection apparatus according to the present invention, the moving object detection section decides, when the total code length of bit plane information inside the region is equal to or smaller than a predetermined second value, that the region is a contour region of the moving object.

[0163] According to this configuration, since the object contours are lines, when a certain region includes too many horizontal direction components, vertical direction components and diagonal direction components, it is possible to determine a region including, for example, a stripe pattern instead of contours of the moving object, and to therefore prevent erroneous detections.

[0164] Furthermore, in the moving object detection apparatus according to the present invention, the motion information extraction section extracts a motion vector from a region decided to be the contour region of the moving object and the moving object detection section decides, when the magnitude of the motion vector is equal to or greater than a predetermined third value, that the region is a contour region of the moving object.

[0165] According to this configuration, it is possible to decide that an immobile object is not a moving object

and thereby improve the accuracy of detecting a moving object.

[0166] Furthermore, in the moving object detection apparatus according to the present invention, the motion information extraction section extracts a first motion vector from the region decided to be the contour region of the moving object, selects a region in the neighborhood of the region, extracts a second motion vector from the selected region, the moving object detection section calculates the magnitude of a difference vector between the first motion vector and the second motion vector as a measured value and decides, when the measured value is equal to or smaller than a predetermined fourth value, that the selected region is an internal region of the moving object.

[0167] According to this configuration, since the contour region of the moving object in the video has a speed different from that in the surrounding region, the region other than the contours of the moving object is decided not to be the region of the moving object and it is possible to thereby improve the accuracy of detecting a moving object.

[0168] Furthermore, in the moving object detection apparatus according to the present invention, the motion information extraction section selects a plurality of regions, extracts motion vectors from the respective selected regions and the moving object detection section

determines the magnitude of the difference vector between the first motion vector and the motion vector of the selected region for each selected region and calculates the total value of the magnitudes of the difference vectors of the selected region as the measured value.

[0169] According to this configuration, since the contour region of the moving object in the video has a speed different from that of the surrounding region, it is possible to decide that a plurality of regions other than the contours of the moving object are not the regions of the moving object and improve the accuracy of detecting a moving object.

[0170] Furthermore, in the moving object detection apparatus according to the present invention, the moving object detection section decides, when the magnitude of the difference vector between the motion vector in the region decided to be an internal region of the moving object and the motion vector of the region in the neighborhood of the region is equal to or smaller than a predetermined fifth value, that the region is an internal region of the region of the moving object.

[0171] According to this configuration, it is possible to detect a region of a moving object moving at a certain speed in which the object is not decided to be a moving object and improve the accuracy of detecting a moving object.

[0172] Furthermore, in the moving object detection

apparatus according to the present invention, the moving object detection section decides that a region surrounded by the region decided to be the contour region of the moving object or the internal region of the moving object is an internal region of the moving object.

[0173] According to this configuration, it is possible to detect the inside of the moving object decided to be the contours as the region of the moving object and improve the accuracy of detecting the moving object.

[0174] Furthermore, in the moving object detection apparatus according to the present invention, when the number of regions decided to be a contour region or internal region of a second moving object in the neighborhood of the contour region or internal region decided to be a first moving object equals or exceeds a predetermined sixth value, the moving object detection section re-decides the contour region or internal region decided to be the first moving object as the first moving object.

[0175] According to this configuration, it is possible to decide that too small a region is not a moving object and thereby reduce erroneous detections in moving object detection.

[0176] The moving object detection method according to the present invention is a method for detecting a moving object from a video stream, including a step of extracting motion information from a video stream video-coded using layered coding whereby a video is coded with being divided

into a plurality of layers and motion predictive compensation coding, a step of extracting edge information from the video stream and a step of detecting a moving object using the extracted motion information and the edge information, the steps being executed by the moving object detection apparatus that detects the moving object.

[0177] According to this method, it is possible to detect contours of an object without decoding any video stream, detect a moving object from motion information and detect a moving object at a high speed, with high accuracy and low processing load.

[0178] The moving object detection program according to the present invention is intended to detect a moving object from a video stream by causing a computer to execute a step of extracting motion information from a video stream video-coded using layered coding whereby a video is coded with being divided into a plurality of layers and motion predictive compensation coding, a step of extracting edge information from the video stream and a step of detecting a moving object using the extracted motion information and the edge information.

[0179] According to this program, it is possible to detect contours of an object without decoding any video stream, detect a moving object from motion information and detect a moving object at a high speed, with high accuracy and low processing load.

[0180] The video decoding apparatus according to the present invention includes a video decoding section that decodes a video stream coded by layered coding whereby a video is coded with being divided into a plurality of layers and motion predictive compensation coding and a moving object detection section that detects a moving object from motion information and edge information extracted when the video decoding section decodes the video stream.

[0181] According to this configuration, the video decoding apparatus and moving object detection apparatus can share some processing and section, perform video decoding and moving object detection simultaneously and at a high speed, and reduce the overall scale of the apparatus.

[0182] Furthermore, in the video decoding apparatus of the present invention, the video stream is divided into a plurality of regions and when the total code length of bit plane information inside the region is equal to or greater than a predetermined first value, the moving object detection section decides that the region is a contour region of the moving object.

[0183] According to this configuration, only by confirming the amount of code of bit planes up to a threshold bit position of a certain region of the horizontal direction component, vertical direction component and diagonal direction component, it is

possible to decide the number of edges which exist within the region and detect the contours of the object at a high speed.

[0184] Furthermore, in the video decoding apparatus of the present invention, when the total code length of bit plane information in the region is equal to or smaller than a predetermined second value, the moving object detection section decides that the region is a contour region of the moving object.

[0185] According to this configuration, since the object contours are lines, when a certain region includes too many horizontal direction components, vertical direction components and diagonal direction components, it is possible to determine a region including, for example, a stripe pattern instead of contours of the moving object, and to therefore prevent erroneous detections.

[0186] In the video decoding apparatus of the present invention, the video decoding section generates a video emphasizing the moving object detected by the moving object detection section.

[0187] According to this method, the supervisor can easily detect the moving object.

[0188] In the video decoding apparatus of the present invention, the video decoding section generates a video made up of an edge component and emphasizes and displays only the region of the moving object detected by the moving object detection section.

[0189] In this way, even when the bit rate of the basic layer is very low due to restrictions on the communication speed and only videos of extremely bad image quality can be generated, contours alone may rather help recognition of details.

[0190] Furthermore, in the video made up of contours, only the moving object is quite noticeable and it is easier for the supervisor who monitors a plurality of monitoring videos simultaneously to detect an abnormality or suspicious figure, or in an environment in which a processing capacity is limited, for example, when a plurality of camera videos are displayed, it is possible to make it easier to see important regions from the standpoint of monitoring with low processing load.

[0191] The video coding apparatus of the present invention includes a video coding section that generates a video stream coded using layered coding whereby a video is coded with being divided into a plurality of layers and motion predictive compensation coding and a moving object detection section that extracts motion information and edge information of the video when the video coding section codes the video and detects a moving object. According to this configuration, the video coding section and moving object detection section can share some processing or section, perform video coding and moving object detection simultaneously and at a high speed, and reduce the overall scale of the apparatus.

[0192] The image pickup apparatus of the present invention includes an image pickup section that inputs a video, the video coding apparatus according to the present invention that codes a video input by this image pickup section, an image pickup control section that controls an image pickup function for the image pickup section based on a moving object detection result output by the moving object detection section and an output section that outputs the video stream and the detection result of the moving object.

[0193] This configuration makes it possible to detect a moving object in the process of generating a video stream for video transmission to a remote place and therefore to continue to detect, take images of a suspicious figure or the like as a moving object at a high speed during video monitoring or the like, and to transmit the video and efficiently perform video monitoring.

[0194] Furthermore, in the image pickup apparatus of the present invention, the image pickup control section controls the image pickup section so that the area of the region of the moving object output by the moving object detection section is kept to a constant proportion with respect to the total area of the input video.

[0195] This configuration makes it possible to include the moving object and its surrounding situation in the video and achieve efficient monitoring of a focused moving object.

[0196] The video monitoring system of the present invention includes the image pickup apparatus according to the present invention and a video monitoring apparatus that decodes the video stream received from this image pickup apparatus and recognizes the image in the region of the detected moving object using the detection result of the moving object.

[0197] This configuration allows a moving object to be detected in the process of generating a video stream for video transmission to a remote place, making it possible to omit image recognition processing of regions other than the moving object and perform image recognition at a high speed and with low processing load, thereby continue to detect and take images of a suspicious figure or the like at a high speed as a moving object during video monitoring.

[0198] Note that image recognition in the present invention is not limited to detection of a moving object, but it refers to an automatic mechanical decision section using an image including recognition of a figure, face, object or personal authentication.

[0199] Furthermore, in the video decoding apparatus of the present invention, the video stream is coded with being layered into a basic layer and expanded layer, the motion information extraction section extracts the motion information from the video stream of the basic layer and the edge information extraction section extracts the edge

information from the video stream of the expanded layer.

[0200] According to this configuration, when the motion information indicates that there is no motion, it is possible to stop processing such as extraction of the edge information and reduce the processing load, and when the edge information indicates that there is no edge, it is possible to stop processing such as extraction of the motion information and reduce the processing load and thereby detect contours of an object at a high speed.

[0201] Furthermore, in the video decoding apparatus of the present invention, the video stream is coded with being layered into a basic layer and expanded layer, the motion information extraction section extracts the motion information from the video stream of the expanded layer and the edge information extraction section extracts the edge information from the video stream of the expanded layer.

[0202] According to this configuration, it is possible to perform detection processing of the moving object using only the video stream of the expanded layer and detect contours of the object at a high speed and with fewer video streams.

[0203] The present application is based on Japanese Patent Application No.2004-161053 filed on May 31, 2004 and Japanese Patent Application No.2005-035627 filed on February 14, 2005, entire content of which is expressly

incorporated by reference herein.

Industrial Applicability

[0204] The present invention is suitable for use in a moving object detection apparatus that detects a moving object from a video stream generated by coding a video and suitable for detecting a moving object at a high speed without decoding a video stream.